基于生成式对抗网络的中文字体风格迁移 *

滕少华，孔棱睿

(广东工业大学 计算机学院，广州 510006)

摘　要：风格多样的中文字体是一种重要的中国文化符号，它的设计和操作是一项需要大量专业知识的艰巨工作。因此，针对这项工作提出一种基于生成式对抗网络的中文字体风格迁移的新方法。实验中，使用基于残差网络结构的生成式模型，在均方误差约束下，进行生成式模型与判别式模型之间的对抗训练，最后使用训练所得的生成式模型实现不同中文字体间一对一和多对多的风格迁移。实验表明，与之前常用的基于 L1 正则化方法相比，使用这种方法在字体细节生成上有更出色的表现，简化了中文字体的建模方式，提高了生成图像的逼真度，并具有更好的灵活性和通用性。

关键词：风格迁移；生成式对抗网络；卷积神经网络；残差网络；深度学习

中图分类号：TP391.12　　doi: 10.3969/j.issn.1001-3695.2018.04.0330

Chinese fonts style transfer based on generative adversarial networks

Teng Shaohua, Kong Lengrui

(*College of Computer Guangdong*, *University of Technology*, *Guangzhou* 510006, *China*)

Abstract: A variety of Chinese characters is an important Chinese cultural symbol. Its design and operation is a hard work requiring a lot of professional knowledge. Therefore, for this work, this paper proposed a new method of Chinese font style transfer based on Generative Adversarial Networks. In the experiment, using a generative model based on the residual network structure, performing the adversarial training between the generative model and the discriminative model under the constraint of mean square error. At last, the trained generative model could be used to implement one-to-one and many-to-many style transfer between different Chinese fonts. Experiments show that, compared with the usual L1 regularization method used before, the method proposed in this paper has better performance in font detail generation, simplified Chinese font modeling, improved the fidelity of generated images, and has better flexibility and versatility.

Key words: style transfer; generative adversarial networks(GANs); convolutional neural networks; residual networks; deep learning

0　引言

中文字体是一种二维设计的视觉元素，它是汉字文化的主要表现形式。与英文字体相比，博大精深的汉字具有字义丰富和形体复杂的特点。文字专家需要花费大量的时间来设计不同风格的中文字体，这是一项劳动密集型的工作。而且中文字体的种类繁多，往往中文字体库所含的字体风格不能满足实际需求。因此，本文提出了一个用于自动转换字体风格的神经网络架构，可将所需的字体实例应用到目标的项目中。

对于字形的早期研究主要基于几何建模，只能用于特定的字形拓扑结构，不能用于图像输入。Shamir 等人[错误!未找到引用源。]提出了一种基于参数化特征的字体设计的方法，但是这种方法的特征和约束的提取过程不是自动完成的，需要用户干涉。Suveeranont 等人[错误!未找到引用源。]提出了一种从用户定义的实例中自动生成新字体的方法，可以从每个字体轮廓中提取拓扑结构，然后通过特征保留和加权混合来生成新字体，然而所生成的新字体容易出现明显的扭曲。与简单的英文字体相比，中文字体的风格迁移特别具有挑战性，因为中文字体要比英文字体更复杂多变。曾理等人[错误!未找到引用源。]提出了一种基于多尺度非冗余小波纹理分析的文字种类自动识别技术，可用于识别不同格式和字体的文本图像。Xu 等人[错误!未找到引用源。]提出了一种通过分解和重组字符组件来模仿手写文字风格的智能算法，该算法的文字生

收稿日期：2018-04-04；修回日期：2018-05-25　　基金项目：国家自然科学基金资助项目（61402118，61673123，61603100，61702110）；广东省科技计划项目（2015B090901016，2016B010108007）；广东省教育厅项目（粤教高函［2018］1 号；粤教高函［2015］133 号；粤教高函［2014］97 号）；广州市科技计划项目（201604020145，201604030034，201508010067，201604046017）

作者简介：滕少华（1962-），男，江西省人，教授，博士，主要研究方向为协同计算、大数据、数据挖掘、网络安全；孔棱睿（1986-），男，广东英德人，硕士，主要研究方向为机器学习、深度学习、计算机视觉（konglengrui@126.com）.

成过程过于复杂繁琐。陈飞等人[错误!未找到引用源。]提出了基于 SDL 和 OpenGL 实时绘制中文字体的方法，优化了中文字体生成纹理图的过程；但是该方法过度依然于图形库，不利于功能上的扩展。

随着深度学习的发展[错误!未找到引用源。,错误!未找到引用源。]，研究人员开始对图像中的字形建模展开研究。Baluja [错误!未找到引用源。]在一小组字母上成功地使用深度学习模型来区分字体，并生成相同风格的字符。Azadi 等人[错误!未找到引用源。]提出了一个端到端的堆栈条件的生成式对抗网络模型，有效地实现了 10 000 种不同英文字母字体间的风格迁移。Zhang 等人[错误!未找到引用源。]提出了一种使用递归神经网络（RNN）分别作为识别汉字的判别式模型和生成汉字的生成式模型的框架，可是对于字形复杂的字体生成效果并不理想。Lyu 等人[错误!未找到引用源。]提出了一个可生成中国书法风格字体图像的生成式对抗网络模型；但是该模型设计过于复杂，不利于调整，并且使用了 L1 正则化，容易使得最终生成的字体图像出现笔划粘连或模糊的问题。

本文通过结合条件生成式对抗网络（conditional generative adversarial nets,CGANs）[错误!未找到引用源。错误!未找到引用源。]和 Wasserstein 生成式对抗网络（Wasserstein GAN）[错误!未找到引用源。]来改进中文字形建模的方法。其中，生成式模型是深度学习在计算机视觉领域中一个重要的研究热点。He 等人[错误!未找到引用源。]认为残差网络能有效地识别图像中的抽象特征，并在图像分类中获得了很高的精确度。因此，本文的生成式模型采用灵活性较好的残差网络作为核心组件，并通过条件生成式模型对中文字体的分布进行建模，使得该模型通过从残差网络相关的散度分布中抽样来生成效果逼真的中文字体。

1　生成式对抗网络基本原理

生成式对抗网络由生成式模型和判别式模型组成，如图 1 所示。其中，判别式模型的任务是判断给定的图像是否真实自然，即是否是人为生成的。而生成式模型的任务是生成真实自然的图像，并与原始数据的分布尽可能一致。在训练过程中，生成式模型和判别式模型构成了一个动态的博弈过程，两者在迭代过程中不断优化，而当生成式模型不能区分真实数据和生成数据时，此时可认为生成式模型达到近似最优。下面对本文方法展开探讨。

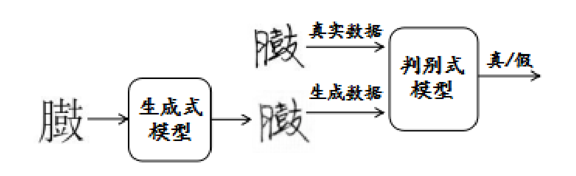


图 1　生成式对抗网络架构

1.1　条件生成式对抗网络

最初的生成式对抗网络（GANs）[错误!未找到引用源。]包含了生成式模型 G 和判别式模型 D 两部分，它从一组先验随机噪声向量开始，通过一个生成式模型与一个判别式模型进行对抗训练，生成一个具有特定分布的图像。在对抗训练过程中，判别式模型试图区分真实图像和假图像，而生成式模型通过试图生成逼真的图像来对抗判别式模型。在理想状态下，判别式模型和生成式模型会处于纳什均衡[错误!未找到引用源。]的状态。通过在生成式对抗网络中的生成式模型和判别式模型添加一些额外信息作为条件，则可以扩展为条件生成式对抗网络，达到控制网络输出的目的，其中额外信息可以是任何类型的辅助信息。在条件生成式对抗网络[错误!未找到引用源。]中，先验随机噪声向量和额外信息以联合形式组合，形成有条件的映射关系。因此，生成式模型与判别式模型之间的对抗训练过程可以用以下损失函数表示：

$$L_{CGAN}(G,D)=E_{x\sim P_r(x)}[\log D(x|c)]+ E_{z\sim P_z(z)}[\log(1-D(G(z|c)))] \quad (1)$$

其中：$P_r(x)$ 表示真实数据；$P_z(z)$ 表示随机噪声数据；c 表示额外信息；生成式模型 G 和判别式模型 D 交替地最小化和最大化这个损失函数最终求得近似最优解的生成式模型 G^*。

1.2　Wasserstein 生成式对抗网络

最初的生成式对抗网络[错误!未找到引用源。]的优化策略是基于 JS 散度（jensen shannon divergence）和 KL 散度（kullback leibler divergence），而 Arjovsky 等人[错误!未找到引用源。]证明了这种优化方式会导致对抗训练过程不稳定，判别式模型与生成式模型之间的均衡问题难以解决，以及容易出现梯度消失和模型崩溃的问题。针对上述问题，Arjovsky 等人[错误!未找到引用源。]提出了使用 EM 距离（earth-mover distance）作为优化策略的 Wasserstein 生成式对抗网络，其中，EM 距离（亦可称为 Wasserstein 距离）的定义如下：

$$W(P_r,P_g)=\min_{r\sim\prod(P_r,P_g)} E_{(x,y)\sim\gamma}[\|x-y\|] \quad (2)$$

其中：$\prod(P_r,P_g)$ 表示边界为 P_r 和 P_g 的所有联合分布 (x,y) 的集合；$\|x-y\|$ 表示每一个可能的联合分布中真实样本 x 和生成样本 y 的距离，则该联合分布 γ 下所有样本对的距离期望值可表示为 $E_{(x,y)\sim\gamma}[\|x-y\|]$。根据 Kantorovich Rubinstein 对偶原理[错误!未找到引用源。]，对式（2）作进一步推导可得出 Wasserstein 生成式对抗网络[错误!未找到引用源。]的表达式如下：

$$W(P_r,P_g)=\max_{|D|_L\le 1} E_{x\sim P_r(x)}[D(x)]- E_{x\sim P_x(x)}[D(x)] \quad (3)$$

其中：$P_r(x)$ 表示真实数据；$P_g(x)$ 表示生成数据；$|D|_L\le 1$ 表示判别式模型 D 是 Lipschitz 常数为 1 的函数。

2　中文字体风格迁移方法

2.1　一对一字体风格迁移方法

基于生成式对抗网络在图像生成式模型的成功建立，通过探索了可用于字体风格迁移的各种生成式对抗网络体系结构，并进行了大量的实验以发现有效稳定的网络结构。在众多体系结构中，本文采用了 Wasserstein GAN 作为模型优化的基本策

略。文献**错误!未找到引用源。**提出之前基于 JS 散度和 KL 散度优化方式的 GAN 需要通过加入噪声数据的方案来克服训练不稳定和梯度消失的问题，而文献**错误!未找到引用源。**中论证了即使两个数据分布没有任何重叠，Wasserstein 距离仍然能够反映出它们的距离。因此，可以仅将字体图像作为输入数据馈送到生成式模型中，然后进行端到端的对抗训练，最终有效地生成目标字体图像。

源字体　　　　风格字体

L1　　　　MSE

图 2　L1 与 MSE 的效果对比

在传统的生成式对抗网络中，生成式模型的目标函数主要与 L1 正则化结合以获得逼真的图像生成效果。但是中文字体风格迁移对生成图像的形状和轮廓的辨识度要求更高，使用 L1 正则化的方法容易导致生成字体图像出现笔划粘连或模糊的问题，而在处理较复杂的中文字体时，此类问题显得更为突出。为了克服这些问题，本文方法使用均方误差(mean-square error ,MSE)作为生成式模型训练的约束条件，如图 2 所示，所生成的字体图像在字形和轮廓上更加清晰，字体辨识度更高。

判别式模型与生成式模型之间的对抗关系可以表述为以下损失函数：

$$L_w(G,D)=E_{x\sim P_r(x)}[D(x)]-E_{y\sim P_g(y)}[D(G(y))] \tag{4}$$

其中：y 表示输入的源字体图像数据；x 表示目标风格的字体图像数据；$G(y)$ 表示生成式模型 G 的输出；判别式模型 D 为满足 Lipschitz 常数为 1 的函数。此时，生成式模型的目标函数 G^* 可以表示为

$$G^*=\arg\min_G\max_{|D|_L\le 1}L_w(G,D)+\lambda L_{MSE}(G) \tag{5}$$

其中：$L_{MSE}(G)=E_{x\sim P_r(x),y\sim P_g(y)}[\|x-G(y)\|_{MSE}]$。

2.2　多对多字体风格迁移方法

一对一字体风格迁移方法采用端到端的训练方式，目标是将输入字体图像转换成具有新风格的字体图像，但是这种一对一的映射关系无法满足多种字体转换多风格的实际应用需求。为满足这个需求，本文在一对一字体风格迁移方法的基础上增加了迁移条件，从而可以控制生成式模型的目标输出，进而实现一对多（1:N）、多对一（N:1）和多对多（N:M）字体风格迁移，其中，N 与 M 为大于 1 的整数。

在多对多字体风格迁移方法中，字体集 A 与 B 分别作为网络模型的输入与输出，它们之间的关联关系有以下四种：

a)一对一（1:1）。当 A 与 B 中都只有 1 种风格字体时，A 与 B 之间属于 1:1 关联。此时，对于 A 中的每一个字体，B 中至多有一个字体与之关联。

b)一对多（1:N）。当 A 只有一种风格字体，B 有 N 种风格字体，A 与 B 属于 1:N 关联。此时，对于 A 中的每一个字体，B 中有 N 个字体与之关联。

c)多对一（N:1）。当 A 有 N 种风格字体，B 只有 1 种风格字体，A 与 B 属于 N:1 关联。此时，对于 B 中的每一个字体，A 中有 N 个字体与之关联。

d)多对多（N:M）。当 A 有 N 种风格字体，B 只有 M 种风格字体，A 与 B 属于 N:M 关联。此时，对于 A 中的每一个字体，B 中有 M 个字体与之关联，而对于 B 中的每一个字体，A 中有 N 个字体与之关联。

本文的网络模型可以单独或混合地实现以上四种关联关系。根据条件生成式对抗网络**错误!未找到引用源。**的基本原理，在式（4）的基础上添加了一个额外信息作为字体风格迁移条件，即输入字体图像与转换条件组合后作为输入馈送到网络模型中。因此，在多对多字体风格迁移方法中，判别式模型与生成式模型之间的对抗关系可以表述为以下损失函数：

$$L_M(G,D)=E_{x\sim P_r(x)}[D(x|t)]-E_{y\sim P_g(y)}[D(G(y|t))] \tag{6}$$

其中：t 表示字体风格迁移的迁移条件；y 表示输入的源字体图像数据；x 表示目标风格的字体图像数据；$G(y|t)$ 表示生成式模型 G 的输出。此时，生成式模型的目标函数 G^* 可以表示为

$$G^*=\arg\min_G\max_{|D|_L\le 1}L_M(G,D)+\lambda L_{MSE}(G) \tag{7}$$

其中：$L_{MSE}(G)=E_{x\sim P_r(x),y\sim P_g(y)}[\|x-G(y|t)\|_{MSE}]$。

3　实验

3.1　实验平台与数据

本文实验中使用的操作系统为 Ubuntu 16.04 操作系统，内存 8 GB，双核 Inter CPU I5，图像处理器（GPU）采用 NVIDIA GTX 1050Ti，以及使用基于 GPU 版本的 Tensorflow1.2 深度学习框架。由于目前还没有用于字体风格迁移实验的公开数据集，所以本文分别从华文宋体、华文行楷和方正静蕾简体这三种风格字体中随机抽取 1 000 个字体组成实验数据集，其中，90%的字体用于字体风格迁移实验，10%的字体用于评估网络模型的泛化性。

3.2　网络模型的实现详情

本文的字体风格迁移网络遵循生成式对抗网络的基本架构。其中，判别式模型是一个二分类器，由卷积神经网络层与 pooling 层结合的网络结构，而生成式模型不使用任何 pooling 层，而是使用步长和步长卷积来进行网络内部的下采样和上采样。

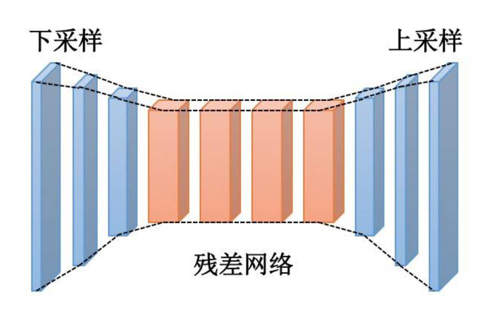


图 3　生成式模型网络结构

中文字体具有多样和复杂的布局，使用残差网络块能够有效地学习字体的空间结构之间的微小差异。如图 3 所示，本文的生成式模型网络体由下采样块、残差网络和上采样块组成。在下采样块部分，一个字符图像将被卷积到 16*512 个特征映射，分别由 4 个卷积层和 64、128、256、512 个滤波器组成。在残差网络部分，由 4 个残差网络块串联而成，输入和输出均为 16*512 特征映射。在上采样块部分，先前的 16*512 特征映射被逐步解卷积，分别由 2 个解卷积层和 2 个全卷积层和 256、128、64、64 个滤波器组成。除了输出层以外，所有网络层都由实例标准化(instance normalization)[错误!未找到引用源。] 和 ReLU 激活函数连接而成，使用 3*3 卷积核，而输出层使用标量化的 tanh 激活函数来确保输出图像的像素取值范围为[0,255]。生成式模型网络的输入和输出均是张量为 3*64*64 的图像数据，下采样和上采样的步长分别为 2 和 1/2。判别式模型由 4 个卷积层和只有一个输出的全连接层组成,使用 3*3 卷积核,每个卷积层之后是批量标准化（batch normalization）和 ReLU 激活函数。

在对抗训练过程中，判别式模型和生成式模型交替进行随机梯度下降，均使用学习率为 0.000 8 的 RMSprop 算法进行网络优化，学习率的衰退系数为 0.8，训练数据批量大小为 16，总共进行 1 000 次迭代。

3.3　一对一字体风格迁移

为了评估本文模型在一对一字体风格迁移中的性能，实验数据由 3 种风格差别较大的字体中各随机抽取 1 000 个字体组成，分别是印刷风格的华文宋体、书法风格的华文行楷以及手写风格的方正静蕾简体。其中，华文宋体字体作为网络模型中的输入，即源字体，华文行楷和方正静蕾简体作为网络模型的目标输出，即风格字体。实验结果如图 4 所示。本文模型在字体风格迁移中达到了字形和轮廓不变的基本要求，图像噪声基本可以忽略，生成的字体图像效果逼真。与常用的基于 L1 正则化方法相比，本文方法具有更好的字体笔划细节和清晰度。

源字体　　风格字体　　L1 正则化方法　　本文方法

图 4　一对一字体风格迁移结果

3.4　多对多字体风格迁移

在一对一字体风格迁移方法中，生成式模型只能用于两种字体单向端到端的风格迁移，而在实际应用中往往要求生成式模型可以对多种字体进行风格迁移，这可以通过引入风格迁移条件来实现。为了简单有效，本文使用单通道的灰度图像表示风格迁移条件，对于不同字体风格间的迁移条件采用数字符号进行表示。例如数字 0 表示从华文行楷风格迁移到方正静蕾简体，数字 1 表示从华文行楷风格迁移到华文宋体，依次类推。此时，网络输入是字体图像和风格迁移条件组合而成的 4*64*64 张量数据，并且无须修改其他网络参数，不会增加额外计算量。实验结果如图 5 所示。使用不同的字体作为网络的输入与输出，并增加了额外的风格迁移条件，网络模型仍然可以生成逼真的字体图像，显示了本文方法具有良好的鲁棒性。

3.5 模型的泛化性评估

在深度学习中，泛化性是模型训练的要点之一。简单地说，泛化性是指在训练数据集上训练好的模型对未知数据的预测能力。如图 6 所示，本文网络模型可以粗略地预测到字体的字形和轮廓，其中部分字体表现良好，显示了本文方法具有一定的泛化性，对新字体的设计具有实际参考价值。

源字体 风格字体 生成结果 迁移条件

0

1

2

3

4

5

图 5　多对多字体风格迁移结果

华文行楷　方正静蕾简体



图 6　预测字体结果

4　结束语

本文在对条件生成式对抗网络和 Wasserstein 生成式对抗网络的网络结构研究基础上，采用灵活性较好的残差网络结构作为生成式模型的核心组件，提出了一种基于生成式对抗网络的中文字体风格迁移的新方法，有效地实现了不同风格字体间的一对一和多对多的风格迁移。实验表明，本文方法适用于中文字体风格迁移，能够生成逼真的字体图像。但是本文方法仍然有一些的局限性，如生成结果图像的字体大小是固定的，要改善这个问题需要更多的时间进行探索，以及收集更多的数据，并将会在未来的工作中考虑。

参考文献：

[1] Shamir A, Rappoport A. Feature-based design of fonts using constraints [C]// Lecture Notes in Computer Science. 1998: 93-108.

[2] Suveeranont R, Igarashi T. Example-based automatic font generation [C]// Proc of International Conference on Smart Graphics. Berlin: Springer-Verlag, 2010, 6133: 127-138.

[3] 曾理，唐远炎，陈廷槐. 基于多尺度小波纹理分析的文字种类自动识别 [J]. 计算机学报, 2000, 23 (7): 699-704 (Zeng Li, Tang Yuanyan, Chen Tinghuai. Multi-scale wavelet texture-based script identification method [J]. Chinese Journal of Computers, 2000, 23 (7): 699-704.) .

[4] Xu S, Jin T, Jiang H, *et al.* Automatic generation of personal Chinese handwriting by capturing the characteristics of personal handwriting [C]// Proc of the 21st Innovative Applications of Artificial Intelligence Conference. [S. I.] : IAAI-09, 2009: 191-196.

[5] 陈飞，黄海明，杨猛，等. 基于 SDL 和 OpenGL 实时绘制中文字体 [J]. 计算机工程与设计, 2011, 32 (8): 2764-2767 (Chen Fei, Huang Haiming, Yang Meng, *et al.* Real-time rendering method of Chinese fonts based on SDL and OpenGL [J]. Computer Engineering and Design, 2011, 32 (8): 2764-2767.) .

[6] 毛勇华，桂小林，李前，等. 深度学习应用技术研究 [J]. 计算机应用研究, 2016, 33 (11): 3201-3205 (Mao Yonghua, Gui Xiaolin, Li Qian, *et al.* Study on application technology of deep learning [J]. Application Research of Computers, 2016, 33 (11): 3201-3205) .

[7] 刘建伟，刘媛，罗雄麟. 深度学习研究进展 [J]. 计算机应用研究, 2014, 31 (7): 1921-1930 (Liu Jianwei, Liu Yuan, Luo Xionglin. Research and development on deep learning [J]. Application Research of Computers, 2014, 31 (7): 1921-1930) .

[8] Baluja S. Learning typographic style [J]. arXiv preprint arXiv: 1603. 04000, 2016.

[9] Azadi S, Fisher M, Kim V, *et al.* Multi-content GAN for few-shot font style transfer [J]. arXiv preprint arXiv: 1712. 00516, 2017.

[10] Zhang X Y, Yin F, Zhang Y M, *et al.* Drawing and recognizing chinese characters with recurrent neural network [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2018, 40 (4): 849-862.

[11] Lyu P, Bai X, Yao C, *et al.* Auto-encoder guided GAN for Chinese calligraphy synthesis [J]. arXiv preprint arXiv: 1706. 08789, 2017.

[12] Mirza M, Osindero S. Conditional generative adversarial nets [J]. Computer Science, 2014: 2672-2680.

[13] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN [J]. arXiv preprint arXiv: 1701. 07875, 2017.

[14] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [J]. arXiv preprint arXiv: 1512. 03385, 2015.

[15] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks [C]// Advances in Neural Information Processing Systems. 2014, 3: 2672-2680.

[16] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks [J]. arXiv preprint arXiv: 1701. 04862, 2017.

[17] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization [J]. arXiv preprint arXiv: 1607. 08022, 2016.

[18] Villani C. Optimal transport: old and new [M]. [S. I.] : Springer Science & Business Media, 2008.